DOI:
10.1111/acps.13066

Document Version
Peer reviewed version

Antipsychotic use and risk of life-threatening medical events: umbrella review of observational studies

Davide Papola[1]; Giovanni Ostuzzi[1]; Chiara Gastaldon[1]; Gian Paolo Morgano[2]; Elena Dragioti[3]; Andre F. Carvalho[4,5]; Paolo Fusar-Poli[6,7,8,9]; Christoph U. Correll[10,11,12,13]; Marco Solmi[6,14,15]*; Corrado Barbui[1]*.

Running head: Antipsychotic use and risk of life-threatening medical events.

[1] WHO Collaborating Centre for Research and Training in Mental Health and Service Evaluation, Department of Neuroscience, Biomedicine and Movement Science, Section of Psychiatry, University of Verona, Verona, Italy;

[2] Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, Canada;

[3] Pain and Rehabilitation centre and Department of Medicine and Health Sciences (IMH), Faculty of Health Sciences University of Linköping, SE- 581 85 Linköping, Sweden;

[4] Department of Psychiatry, University of Toronto, Toronto, ON, Canada;

[5] Centre for Addiction and Mental Health (CAMH), Toronto, ON, Canada;

[6] Early Psychosis: Interventions and Clinical-detection (EPIC) lab, Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK;

[7] OASIS service, South London and Maudsley NHS Foundation Trust, London, UK;

[8] Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy;

[9] National Institute for Health Research, Maudsley Biomedical Research Centre, South London and Maudsley NHS Foundation Trust, London, UK;

[10] The Zucker Hillside Hospital, Department of Psychiatry, Northwell Health, Glen Oaks, NY, USA;

[11] Zucker School of Medicine at Hofstra/Northwell, Department of Psychiatry and Molecular Medicine, Hempstead, NY, USA;

[12] The Feinstein Institute for Medical Research, Center for Psychiatric Neuroscience, Manhasset, NY, USA;

[13] Charité Universitätsmedizin Berlin, Department of Child and Adolescent Psychiatry, Berlin, Germany;

[14] Neuroscience Centre, University of Padua, Padua, Italy;

[15] Department of Neuroscience, University of Padua, Padua, Italy;

***** Joint last authors

Correspondence to: Dr. Davide Papola, MD

WHO Collaborating Centre for Research and Training in Mental Health and Service Evaluation, Department of Neuroscience, Biomedicine and Movement Science,

Section of Psychiatry, University of Verona,

Policlinico Giambattista Rossi, Piazzale L.A. Scuro 10, 37134 Verona, Italy.

Tel +39-045-8124063

E-mail address: davide.papola@univr.it

Summations

- Systematic reviews of observational studies showed a possible increased risk of life-threatening medical events associated with exposure to antipsychotic drugs, however the certainty of the risk estimates has never been quantitatively assessed;
- An umbrella review of the data on the association between antipsychotics and the risk of hip fracture, venous thromboembolism, stroke, myocardial infarction, pneumonia, and sudden cardiac death was conducted applying AMSTAR-2, GRADE and quantitative umbrella review criteria;
- We graded the association between antipsychotic exposure and pneumonia as "convincing", followed by the association on hip fracture and venous thromboembolism. The data on stroke, sudden cardiac death, and myocardial infarction was the least reliable.

Limitations

- Considering the observational nature of the primary studies, confounding by indication may have inflated the risk estimates;
- We found significant heterogeneity in terms of populations included in the primary studies;
- We were unable to re-analyse the data by antipsychotic class or by individual drug, as this information was seldom reported.

Abstract

Objective: To quantify the risk of hip fracture, thromboembolism, stroke, myocardial infarction,pneumonia, and sudden cardiac death associated with exposure to antipsychotics.

Methods: Systematic searches were conducted in Medline, Embase, PsychINFO from inception until 30/07/2018 for systematic reviews of observational studies. AMSTAR-2 was used for quality assessment of systematic reviews, while the strength of associations was measured using GRADE and quantitative umbrella review criteria(URC).

Results: Sixty-eight observational studies from six systematic reviews were included. The association between antipsychotic exposure and pneumonia was the strongest (URC=class I; GRADE=low quality; odds ratio [OR]=1.84, 95% confidence interval [CI]=1.62-2.09; participants=28,726; age=76.2±12.3years), followed by the association with hip fracture (URC=class II; GRADE=low quality; OR=1.57,95%CI=1.42-1.74; participants=5,288,118; age=55.4±12.5years), and thromboembolism (URC=class II; GRADE=very low quality; OR=1.55,95%CI=1.31-1.83; participants=31,417,175; age=55.5±3.2years). The association was weak for stroke (URC=class III; GRADE=very low quality; OR=1.45,95%CI=1.24-1.70; participants=65,700; age=68.7±13.8years), sudden cardiac death (URC=class III; GRADE=very low quality; OR=2.24,95%CI=1.45-3.46; participants=77,488; age=52.2±6.2years), and myocardial infarction (URC=class III; GRADE=very low quality; OR=2.21,95%CI=1.41-3.46; participants=399,868; age=74.1±9.3years).

Conclusion: The most robust results were found for the risk of pneumonia, followed by the risk of hip fracture and thromboembolism. For stroke, sudden cardiac death, and myocardial infarction the strength of association was weak. The observational nature of the primary studies may represent a source of bias.

Abstract word count: 200 words.

INTRODUCTION

Antipsychotics (APs) are prescribed in a wide range of health care settings, including psychiatric care, primary care, and other medical settings. In the United Kingdom, in 2010, olanzapine, quetiapine and, risperidone accounted for 24%, 23% and 17% of all prescriptions of psychotropic drugs, respectively (1). APs are primarily indicated for the treatment of schizophrenia and mood disorder; however, APs are also frequently used off-label for the pharmacological treatment of a wide range of other psychiatric and medical conditions, such as anxiety disorders, insomnia, agitation and dementias (2, 3). In clinical practice, the benefits of APs are often limited by side-effects, including extrapyramidal, metabolic, cardiovascular, hepatic and hematological drug reactions (4). In addition, APs may be associated with an increased risk of life-threatening medical events, such as hip fracture, venous thromboembolism, stroke, myocardial infarction, pneumonia, and sudden cardiac death, especially in older adults (5). While for common side-effects, typically occurring soon after APs are prescribed, randomized clinical trials (RCTs) and systematic reviews (SRs) have been able to precisely describe and quantify the increased risk, for life-threatening medical events RCTs may not provide satisfactory information (6). Therefore, in recent years, several observational studies and subsequently SRs have attempted to describe and quantify the association between AP exposure and life-threatening medical events. However, the available data are rather controversial, fragmented, and difficult to be stratified into a pragmatic risk quantification (7), considering each major medical event in comparison to others. Furthermore, the quality of meta-analyses, which is another challenging issue as it

clearly has an influence on the certainty of the risk estimates, has never been formally synthesized.

Aims of the study

In this project, we quantified the risk of the following six life-threatening medical events associated with AP exposure: hip fracture, venous thromboembolism, stroke, myocardial infarction, pneumonia, and sudden cardiac death. Risks were quantified together with a formal assessment of the certainty of estimates using two methods: the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach, and quantitative umbrella review criteria (7).

METHODS

Study design

With the increased number of systematic reviews available, a logical and appropriate next step has been the conduct of reviews of existing systematic reviews, allowing the findings of separate reviews to be compared and contrasted, thereby providing decision makers in healthcare with the evidence they need (11). For this reason we performed an umbrella review. The principle reason for the conduct of an umbrella review is to summarize the evidence from multiple research syntheses. Basically, umbrella reviews are reviews of previously published systematic reviews and meta-analyses, and consist in the repetition of the meta-analyses following a uniform approach for all factors to allow their comparison (12). Conduct of an umbrella review offers the possibility of addressing a vast array of issues related to a topic of interest.

The conduction of an umbrella review is also ideal to provide a clear picture of broad healthcare areas, to highlight whether the evidence base on a topic is consistent or contradictory, and to explore potential sources of heterogeneity for the findings (11, 13). Umbrella reviews are considered amongst the highest level of evidence (14) and are particularly useful to inform policy and clinical decisions (15). A review protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO), Registration Number: CRD42018083965.

Literature search

Two researchers (DP and GO) independently searched Medline, Embase, PsychINFO, CINAHL, and Epistemonikos from database inception until 30th July 2018 for SRs with quantitative synthesis (meta-analyses) of observational studies investigating the association between AP exposure and risk of hip fracture, venous thromboembolism, stroke, myocardial infarction, pneumonia and sudden cardiac death. The following search algorithm was used: "(antipsychotic) AND (stroke OR sudden cardiac death OR venous thromboembolism OR myocardial infarction OR hip fracture OR pneumonia) AND (meta-analysis OR systematic review)". A predefined search strategy was used (Supporting information: Table S1). No date or language restrictions, or restrictions on the patient population, were applied. Electronic database searches were supplemented by a manual search of reference lists from relevant studies. The titles, abstracts, and full texts of the resulting articles were examined in detail for eligibility, and in case of disagreements, a third author (CB) adjudicated the decision. We

documented included and excluded studies following the Preferred Reporting Items for Systematic Reviews and Meta-analyses reporting standards (PRISMA) (16).

Eligibility criteria

Reviews were considered eligible if the authors had performed a systematic search to identify pertinent studies and performed a meta-analysis. We considered SRs of observational studies (both case-control and prospective/retrospective cohort studies) that examined the association between AP exposure and the risk of developing one of the six medical events described above. We excluded SRs that did not present study-level data, such as for example relative risks (RR) or odds ratios (OR) with 95% confidence intervals (CI). When more than one SR on the same research question was available, the SR with the largest number of component studies that provided study-level effect sizes (ES) was considered for inclusion (17, 18).

Data extraction

From each included SR, two investigators (DP and GO) independently extracted information on first author, year of publication, outcomes, number of included studies, and reported summary meta-analytic estimates. All primary observational studies included in each SR were retrieved and carefully inspected by two members (DP, GO) of the research team. The following information was extracted: year of publication, outcome, criteria used to define the occurrence of the medical event under scrutiny, mean age, sex, number of events of interest (serious adverse event), number of AP exposed and non-exposed subjects, study-specific risk estimates adjusted to the largest number of potential confounders (RR, OR, hazard ratio [HR], or standardised

incidence ratio [SIR]), and the corresponding 95% CI, studied population, study design (case-control or cohort).

We gave priority to data informing on “current use” of “any antipsychotic drug” (drugs belonging to the ATC/DDD Index N05A) (19) at the index date. There was no restriction regarding dosage or route of administration. When the original SR presented study results separately by sex or drug class (e.g., first- and second-generation APs), we combined the summary effects using random effects methods, and then the overall meta-analysis was performed.

Reporting quality of included meta-analyses

The quality of included SRs was independently assessed by two reviewers (GO, CG) using AMSTAR-2 (A Measurement Tool to Assess Systematic Reviews), a 16-point assessment tool of the methodological quality of SRs. AMSTAR2 has good inter-rater agreement, test-retest reliability, and content validity (20). It assesses reviews on the following categories: (1) formulation of the research question; (2) a priori design provided; (3) explanation for the chosen study design of the included studies; (4) comprehensive literature search; (5) study selection; (6) data extraction; (7) presence of a list of excluded studies, along with reason for exclusion; (8) comprehensive description of the main features of the included studies; (9) risk of bias assessment; (10) information about the sources of funding for the studies included in the review; (11) methods for statistical combination of results; (12) assessment of the potential impact of risk of bias of individual studies on the meta-analysis result; (13) discussion / interpretation of the potential impact of risk of bias of individual studies on the meta-analysis result; (14) discussion of the heterogeneity observed in the study results; (15)

likelihood of publication bias; (16) declaration of study authors' conflict of interest. Of these 16 domains, seven can particularly affect the validity of the review and its conclusion and are considered "critical domains" (domains 2 – 4 – 7 – 9 – 11 – 13 – 15). Each item allows for the following response options: yes, partial yes, or no. AMSTAR 2 is not intended to be scored. AMSTAR-2 proposes a scheme for interpreting weaknesses detected in critical and non-critical items: "high quality" studies show no or one non-critical weakness; "moderate" quality studies show more than one non-critical weakness but no critical flaws; "low" quality studies show one critical flaw with or without non-critical weaknesses; "critically low" quality studies show more than one critical flaw with or without non-critical weaknesses (Supporting information: Box S1) (20).

Statistical analysis

For each medical event of interest, we re-estimated the summary ES and its 95% CI using random-effects models because we were expecting high heterogeneity (21). We also estimated the 95% prediction interval (PI) for the summary random-effects estimates. PIs further account for heterogeneity between studies and specify the uncertainty for the effect that would be expected in a new study examining that same research question (22). Heterogeneity was evaluated with Cochran's Q statistic (statistically significant for p-value < 0.10) and quantified with the I^2 metric (23). I^2 ranges between 0% and 100%, and it is considered low, moderate, large and very large for values <25%, 25–49%, 50–74% and >75%, respectively. Egger's test was used to evaluate potential publication and small-study effects biases (24, 25). In particular, a p-value ≤ 0.10 in the regression asymmetry test with a more conservative effect in the

largest study was considered evidence for small-study effects bias. We further evaluated the excess significance, which is a test that examines whether the observed number of studies (O) with statistically significant results (positive studies, $p<0.05$) in each meta-analysis is larger than their expected number (E) (26). For each meta-analysis, E is calculated as the sum of the statistical power estimates for each study in the meta-analysis. The power of each study was calculated by an algorithm using a non-central t distribution (27). The estimated power depends on the plausible ES. As the true ES for any meta-analysis is unknown, we assumed that the most plausible effect is given by the largest study. Excess significance for each meta-analysis was claimed at p-value ≤ 0.10 level (26).

Based on these calculations, we applied quantitative umbrella review criteria (7) to classify the strength of each association as "convincing", "highly suggestive", "suggestive", or "weak" (18, 28-31) (Supporting information: Box S2). Specifically, meta-analyses were free from biases (Class I) if they met the following criteria: p-value $< 10^{-6}$ based on random effects meta-analysis; >1000 cases; low or moderate between-study heterogeneity ($I^2 < 50\%$); 95% PI that excluded the null value; no evidence of small-study effects and excess significance. Highly suggestive association (Class II) criteria required >1000 cases, highly significant summary associations (p-value $< 10^{-6}$ by random-effects) and 95% PI not including the null value. Suggestive evidence (Class III) criteria required only >1000 cases and p-value ≤ 0.001 by random-effects. Weak association (Class IV) criteria required only p-value ≤ 0.05. Associations were considered non-significant if $p > 0.05$. The statistical analysis and the power

calculations were performed using STATA version 12.0 (STATA Corp, College Station, TX, USA). P values were all two-tailed.

In addition to these umbrella review criteria, the overall certainty in the estimates was qualitatively assessed by two reviewers (GO, CG) - with one author (CB) adjudicating the decision in case of discrepancies - using the GRADE method (32). GRADE allows to rate the certainty of estimate for each outcome and supplies a tabular overview of findings easily understandable for patients, policy makers, research planners, guideline developers, and other interested stakeholders (33). Summary of Findings tables were developed using the GRADEProGDT app. According to the GRADE method, the following factors were considered for each outcome of interest: study design, risk of bias, consistency, precision, directness, presence of large effect, dose–response gradient, and publication bias assessed using visual inspection of funnel plots and Egger's regression test (25). We strictly followed the GRADE method for all ratings except for consistency, as for consistency we relied on visual inspection of the forest plots only, and we did not use the I^2 statistics. This choice was made because meta-analyses of observational studies include extremely large sample sizes and therefore estimates are very precise (narrow CIs) leading to artificially high I^2 values (34). Based on GRADE assessments, the certainty of estimates was categorized into high, moderate, low or very low (35) (Supporting information: Box S3).

Sensitivity analysis and meta-regression

Sensitivity analyses were performed to assess whether the strength of associations varied when only the following studies were retained in the analysis: studies that

adjusted for at least 4 covariates; studies with a cohort design. We chose to set 4 covariates as a threshold because we considered as less credible not only unadjusted studies but also studies with scarce adjustment (typically age, sex, body mass index). Unrestricted maximum likelihood random effects meta-regressions were used to investigate whether there was a relationship between age/sex and risk of each of the six considered adverse outcomes. Meta-regressions were performed with Comprehensive Meta-Analysis version 2.

RESULTS

Description of studies included in the meta-analyses

The systematic search yielded 1584 records. After duplicate removal and title and abstract screening, 114 full-text articles were retrieved. Of these, six SRs (36-41), including 68 primary studies, met the umbrella review inclusion criteria (Figure 1; Supporting information: Table S2). Included SRs assessed whether exposure to APs increased the occurrence of hip fracture, venous thromboembolism, stroke, myocardial infarction, pneumonia, and sudden cardiac death. Table 1 summarises the main review and individual study characteristics. There were between 6 and 24 study estimates combined per meta-analysis. All included meta-analyses included >1000 cases, ranging from 4,906 to 159,283. Of the 68 primary studies, 26 (38%) were cohort studies and 42 (62%) adopted a case-control design; 64 studies (94%) provided adjusted risk estimates, and 23 (34%) studies adjusted for <4 covariates, usually age and sex. Study participants were of either sex, with miscellaneous conditions (69.1%) or with dementia (14.7%) or psychiatric conditions (16.2%). Altogether, 58.2 ± 3.6%

were female and the mean age of the meta-analysed samples was 55.7 ± 6 years (miscellaneous conditions = 68.3 ± 8.4; psychiatric conditions = 49 ± 5.3; dementia = 78,5 ± 12.4). The mean age of patients in the studies assessing the risk of AP use for hip fracture was 55.4 ± 12.5, for venous thromboembolism 55.5 ± 3.2, stroke 68.7 ± 13.8, myocardial infarction 74.1 ± 9.3, pneumonia 76.2 ± 13.1, and sudden cardiac death 52.2 ± 6.2.

Quality assessment of the included systematic reviews

Of the six reviews, one was of high quality according to the AMSTAR-2 scoring system (36), one was of moderate quality (41), three received a low quality rating (37, 38, 40), and one was of critically low quality (39) (Table 2). AMSTAR-2 detected that in three of six reviews a study protocol was not available, and study selection criteria were unclear. In addition, a list of excluded studies was not provided in four reviews, and the source of funding for the studies included in the reviews was never reported.

Summary effect size

All six meta-analyses showed significant summary random effects estimates, implying that AP use increased the risk of all life-threatening medical events (Figure 2). However, when using $p \leq 10^{-6}$ as a threshold for significance, the meta-analyses on risk of myocardial infarction, sudden cardiac death, and stroke did not produce significant summary results (Table 3).

The effect of the largest primary study included in the meta-analysis of hip fracture, venous thromboembolism, stroke, myocardial infarction, pneumonia, and sudden cardiac death was significant at $p \leq 0.05$, suggesting increased risk (Table 3). For myocardial infarction, however, the effect of the largest primary study failed to reach

statistical significance. In addition, the effects of the largest studies were more conservative than the summary effects of the meta-analysis in four of the six meta-analyses, indicating that the pooled summary measures were influenced by small studies. When we calculated 95% prediction intervals, in only two meta-analyses the null value was excluded (for pneumonia and hip fracture).

Heterogeneity between studies

There was high heterogeneity (I^2 > 75%) in five meta-analyses and low heterogeneity (I^2 < 50%) in the meta-analysis investigating the association between AP exposure and risk of pneumonia (Table 3).

Small-study effects

Small-study effects were found only in the meta-analysis on AP exposure and risk of venous thromboembolism, suggesting that studies not highlighting an association (i.e. the "negative studies") might exist and might have not been included in the analysis.

Excess significance

No meta-analysis provided evidence of excess significance bias using the largest study estimate as the plausible ES ($p < 0.10$).

Umbrella review criteria

According to the umbrella review criteria, one association (between APs and pneumonia) was "convincing" (class I); two associations (between APs and hip fracture and venous thromboembolism) were "highly suggestive" (class II); three associations (between APs and: stroke, myocardial infarction, sudden cardiac death) were "suggestive" (class III) (Table 3; Supporting information: Content S1). We found no non-significant associations.

Certainty of evidence according to the GRADE approach

Table 4 shows the certainty in estimate for each medical outcome, assessed using GRADE. For the outcomes stroke, venous thromboembolism, sudden cardiac death, and myocardial infarction, the certainty was rated as “very low”, mainly because of inconsistency, imprecision and risk of publication bias. For hip fracture and pneumonia, the certainty was rated as “low”, based on the GRADE baseline assumption of low certainty for observational studies.

Overall ranking

Figure 2 presents a ranking of associations based on AMSTAR-2, umbrella review criteria, and GRADE. The association between AP exposure and pneumonia was the most reliable, followed by hip fracture and venous thromboembolism. The association on stroke, sudden cardiac death, and myocardial infarction was weak.

Sensitivity and meta-regression analyses

Limiting the analysis to studies that adjusted for at least 4 confounders, the association between AP exposure and pneumonia, hip fracture, stroke and sudden cardiac death remained “convincing”, highly suggestive and suggestive, respectively. However, the association with venous thromboembolism was downgraded from highly suggestive to suggestive, and the association with myocardial infarction was downgraded from suggestive to weak (Supporting information: Content S2).

Limiting the analysis to cohort studies was not possible for pneumonia, as only one study employed a cohort design. For hip fracture, the strength of association did not change, while the other associations were downgraded by one level (stroke,

myocardial infarction) and by two levels (venous thromboembolism, sudden cardiac death) (Supporting information: Content S3).

Meta-regression analyses did not detect any relationship between age and risk of hip fracture ($p = 0.54$) and sudden cardiac death ($p = 0.17$). By contrast, for venous thromboembolism ($p < 0.01$), stroke ($p = 0.04$), pneumonia ($p = 0.04$) and myocardial infarction ($p = 0.04$) a negative association with age was observed (Supporting information: Content S4). Meta-regression analyses with sex as the moderator variable did not detect any significant relationship with any of the six medical conditions (Supporting information: Content S5).

DISCUSSION

To our knowledge, this is the first umbrella review summarising the data on the association between AP exposure and risk of life-threatening medical events across a broad range of conditions encompassing schizophrenia, dementia, neurological and other medical conditions (Table 1). Considering the quality of existing SRs, and the certainty of estimates measured qualitatively with GRADE and quantitatively with several umbrella review criteria, the most robust results were found for the risk of pneumonia, followed by the risk of hip fracture and thromboembolism. For stroke, sudden cardiac death, and myocardial infarction the strength of association was weak. Sensitivity analyses provided results in line with these findings. In terms of magnitude of risk, the relative risk increased from 45% for stroke to 84% for pneumonia and was more than doubled for myocardial infarction and sudden cardiac death. In absolute terms, these risks correspond to a number needed to harm (NNH) ranging from 87 for

pneumonia to more than 1,800 for sudden cardiac death (Supporting information: Table S3).

For myocardial infarction, pneumonia, venous thromboembolism and stroke, the increased risk associated with APs was higher in studies carried of younger populations versus studies in older populations. This finding appears to be counterintuitive, as the incidence of these outcomes increases with age in the general population, probably because most risk factors, including comorbidities, are age-related (42-45). However, there was a significant difference in the age across study populations, in that studies conducted in patients with a psychiatric diagnosis (schizophrenia, bipolar disorder), who were likely exposed to APs to a relevant higher degree, had much lower mean ages than studies conducted in other patient populations, who were likely exposed to APs only in rare occasions. Moreover, it is possible that in late life, when the absolute baseline risk for the medical life threatening outcomes under study is high, AP exposure adds little relative risk. As the included studies enrolled participants suffering from miscellaneous clinical conditions, subgroup analyses by diagnosis was not possible. However, the review by Papola and colleagues (36) and Nosè and colleagues (40) found no significant differences between patients with psychiatric diagnoses (schizophrenia and bipolar disorder) and patients with cognitive impairment or dementia on the risk of hip fracture and pneumonia, respectively. By contrast, Huang and colleagues (39) reported a higher risk of myocardial infarction in patients with schizophrenia (OR=2.25, CI 1.98 to 2.55, I^2=6%), as compared with those with other diagnoses. Moreover, results from Hsu (38) and colleagues suggested that elderly patients

without dementia might have a higher risk of cardiovascular events than patients with dementia (OR=1.49, CI=1.25–1.77, I^2=89% VS OR=1.17, CI=1.08–1.26, I^2=0%, respectively).

The different strength of association between AP use and medical events have implications for clinical practice. As in clinical practice APs are used in a miscellaneous group of diagnostic entities, which go well beyond the presence of approved indications, such as mania, delusions and hallucinations or augmentation of antidepressants in patients with suboptimal antidepressant response (2, 3), deciding whether an AP should be prescribed should consider these life-threatening medical conditions, which, in turn, may negatively interact with pre-existing physical health conditions. This risk-benefit evaluation may be extremely relevant, for example, for older adults with behavioural and psychological symptoms of dementia, where epidemiological data have documented extensive use of APs, despite evidence of only modest beneficial effects (46). There is also evidence of increasing use of second-generation APs for insomnia and related sleep disorders (47), and high AP prescribing has additionally been documented in nursing home residents (48). In all these circumstances, it would be important to consider if the expected benefit of treatment clearly outweighs the potential drawbacks, especially if treatment is not occasional. Among potential drawbacks, not only the most frequent and well-known AP adverse effects, but also the less frequent, but potentially life-threatening medical conditions need to be considered, especially the risk of pneumonia, where the evidence is most robust. Particular attention needs to be given to the fact that the frequency of these medical conditions is likely to increase in the presence of pre-existing risk factors. For example, obesity, smoking, increasing age, physical inactivity and a sedentary life-style

are all risk factors for thromboembolism (49). Therefore, the different strengths of associations identified in this umbrella review should be used in conjunction with knowledge about risk factors inherent in the patient and associated with co-treatments, among other considerations, to optimize and personalize the individualized choice, use and monitoring of APs in clinical practice.

These different levels of evidence may also have implications for guideline development, in particular for conditions where the use of APs is not backed by substantial evidence of efficacy, such as for example anxiety disorders, insomnia and sleep problems, dementia-related behavioural disinhibition, neurological and medical conditions. In these conditions, tolerability clearly becomes a key argument, and guidelines may therefore consider, in addition to randomised evidence, the differential risk of these rare, but life-threatening medical conditions. By contrast, as in individuals with several severe mental disorders the use of APs is backed by substantial evidence of efficacy, and the prevalence of such serious adverse events is rare, with high NNH as opposed to low number needed to treat (NNT) (50), different recommendations may be drawn based on these results in these circumstances.

A number of limitations associated with this study need to be considered. A general cautionary note is that, although all the statistical tests that we applied are standard in this field (23-27), their use for the URC classification is relatively new, and employed so far to classify strength of associations in some areas only. A second general cautionary note is the observational nature of the primary studies which does not allow to establish a causal association. Very few among the included studies fully matched the two compared populations for underlying risk factors, raising concerns about the

possibility of confounding by indication. The possibility that receiving AP drugs reflected more vulnerability to experiencing the outcomes of interest cannot be ruled out, and therefore the extent to which adjustment for baseline factors accounted for this potential imbalance remains uncertain. It should be noted, however, that this risk is taken into consideration by GRADE, which suggests to rate the certainty of estimates from observational studies as “low quality” instead of “high quality”, to fully acknowledge confounding by indication. Another concern is heterogeneity in terms of populations included in the primary studies. Clinically, it would have been relevant to stratify the risk of medical events by variables that may act as risk factors for the occurrence of these medical outcomes, including medical and psychiatric conditions. Unfortunately, the diagnostic groups were too heterogeneous and sparsely characterized to allow for a moderator analysis based on the primary diagnosis. Indeed, despite attempts to describe the main characteristics of the included populations, in most studies very heterogeneous populations were enrolled, making it very difficult to clearly define to which specific population the results may apply. As for SRs of clinical trial data, future SRs of observational studies should attempt to access individual-level patient data, so that the meta-analytic approach can be applied to more homogeneous subgroups of participants, for example those meeting some diagnostic criteria, or those belonging to a particular age group or setting of care. We argue that the culture of data sharing (51, 52), which has progressively received increasing attention in the field of clinical trials, will be similarly applied to the field of observational studies (53), aiming to make the best use of available data. A second limitation is that we were unable to re-analyse the data by AP class or by individual

drug, as this information was seldom reported. Again, this would have been extremely relevant clinically, especially for those situations where prescribers have already decided that an AP should be prescribed, but the decision of which AP should be selected may be controversial. In these circumstances, information on differences between APs would have been useful to guide clinical practice. In addition to drug class or type, other information that would have been important to re-analyse is whether the risk of medical events is moderated by AP dose and length of treatment. Unfortunately, this was not feasible due to the nature of primary data.

Nevertheless, despite these limitations inherent in the primary data, to our knowledge, this is the first umbrella review that critically appraised the existing data on a topic using an approach based on three complementary tools. AMSTAR-2 was used to assess the reporting quality of SRs, GRADE allowed to qualitatively rate the certainty of estimates by outcome of interest, and quantitative umbrella review criteria were employed to quantitatively measure the strength of associations between AP drug exposure and the outcomes of interest. Based on these tools, it was possible to generate a clinically useful hierarchy for the six considered medical outcomes. We reasoned that GRADE and the umbrella review criteria have complementary characteristics, and therefore their simultaneous use was intended to provide a more complete description of the available data. More specifically, the umbrella review criteria are based on quantitative indicators, such as number of included participants, significance level of statistical associations, evidence of small-study effects and of excess of significance bias, calculation of prediction intervals, presence of heterogeneity, and whether the largest primary study provides significant results.

Being based on objective criteria, it is possible to categorize the strength of associations into different levels, allowing comparisons and informing clinical and policy practice. However, the context to which the results may be applied is not taken into consideration. By contrast, GRADE, which is only partially based on objective measures, takes context aspects into careful consideration. For example, according to the GRADE approach, the extent of similarity (in terms of populations, interventions, outcome measures, and comparisons) between the available studies and the target context of application is a key component of the certainty of estimates. This makes the GRADE approach substantially different from the umbrella review criteria, as it provides a wider perspective that is based on qualitative judgments that account for applicability and context-related aspects.

In conclusion, the strength of association between antipsychotic use and risk of pneumonia was convincing, while the association with hip fracture and thromboembolism was highly suggestive, yet with a low quality according to GRADE. All other associations were weaker. These conclusions should consider that the observational nature of the primary studies cannot prove causality.

Declaration of interest

Dr. Correll has been a consultant and/or advisor to or has received honoraria from: Alkermes, Allergan, Angelini, Boehringer-Ingelheim, Gerson Lehrman Group, Indivior, IntraCellular Therapies, Janssen/J&J, LB Pharma, Lundbeck, MedAvante-ProPhase, Medscape, Merck, Neurocrine, Noven, Otsuka, Pfizer, Rovi, Servier, Sunovion, Supernus, Takeda, and Teva. He has provided expert testimony for Bristol-Myers Squibb, Janssen, and Otsuka. He served on a Data Safety Monitoring Board for Boehringer-Ingelheim, Lundbeck, Rovi, Supernus, and Teva. He received royalties from UpToDate and grant support from Janssen and Takeda. He is also a shareholder of LB Pharma. Dr. Fusar-Poli received grant funds or advisory board fees from Lundbeck.

The other authors have no conflicts of interest to declare.

Funding

This study received no funding.

Authors' contributions

DP and CB conceived and designed the study. DP, GPM, ED, GO, CG, MS and CB carried out the acquisition, analysis or interpretation of data. DP, GO, CC and CB drafted the manuscript, tables, and figures. DP, GO, MS, AFC, PFP and CB performed and refined the statistical analysis. All authors read, edited and approved the final manuscript.

Data availability statement

The data that support the findings of this study are available from “Table 1” and the "Supporting information" file.

Ethics approval and consent to participate

Not applicable.

References

1. ILYAS S, MONCRIEFF J. Trends in prescriptions and costs of drugs for mental disorders in England, 1998-2010. Br J Psychiatry 2012;**200**:393-8.

2. CARTON L, COTTENCIN O, LAPEYRE-MESTRE M, et al. Off-Label Prescribing of Antipsychotics in Adults, Children and Elderly Individuals: A Systematic Review of Recent Prescription Trends. Curr Pharm Des 2015;**21**:3280-97.

3. MAGLIONE M, MAHER AR, HU J, et al. Off-Label Use of Atypical Antipsychotics: An Update. Rockville (MD): Agency for Healthcare Research and Quality (US); 2011.

4. SOLMI M, MURRU A, PACCHIAROTTI I, et al. Safety, tolerability, and risks associated with first- and second-generation antipsychotics: a state-of-the-art clinical review. Ther Clin Risk Manag. 2017;**13**:757-77.

5. JACKSON JW, SCHNEEWEISS S, VANDERWEELE TJ, BLACKER D. Quantifying the role of adverse events in the mortality difference between first and second-generation antipsychotics in older adults: systematic review and meta-synthesis. PLoS One. 2014;**9**:e105376.

6. FRIEDEN TR. Evidence for Health Decision Making - Beyond Randomized, Controlled Trials. N Engl J Med 2017;**377**:465-75.

7. SOLMI M, CORRELL CU, CARVALHO AF, IOANNIDIS JPA. The role of meta-analyses and umbrella reviews in assessing the harms of psychotropic medications: beyond qualitative synthesis. Epidemiol Psychiatr Sci 2018;**27**:537-42.

8. National Institute for Health and Care Excellence (2016) Psychosis and schizophrenia in adults: prevention and management (NICE Guideline 178).

https://www.nice.org.uk/guidance/cg178/chapter/About-this-guideline. Accessed November 15, 2018.

9. Scottish Intercollegiate Guidelines Network (2013) Management of schizophrenia (SIGN Guideline 131). https://www.sign.ac.uk/assets/sign131.pdf. Accessed November 15, 2018.

10. World Health Organization (2012) Antipsychotic medications for psychotic disorders. https://www.who.int/mental_health/mhgap/evidence/psychosis/q1/en/ Accessed November 15, 2018.

11. AROMATARIS E, FERNANDEZ R, GODFREY CM, HOLLY C, KHALIL H, TUNGPUNKOM P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc 2015;**13**:132-40.

12. FUSAR-POLI P, RADUA J. Ten simple rules for conducting umbrella reviews. Evid Based Mental Health 2018;**21**:95-100.

13. IOANNIDIS JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ 2009;**181**:488-93.

14. GRANT MJ, BOOTH A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health information and libraries journal. 2009 Jun;26:91-108.

15. FUSAR-POLI P, HIJAZI Z, STAHL D, STEYERBERG EW. The Science of Prognosis in Psychiatry: A Review. JAMA psychiatry. 2018 Dec 1;75:1289-97.

16. MOHER D, LIBERATI A, TETZLAFF J, ALTMAN DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. BMJ 2009;339.

17. SOLMI M, KOHLER CA, STUBBS B, et al. Environmental risk factors and nonpharmacological and nonsurgical interventions for obesity: An umbrella review of meta-analyses of cohort studies and randomized controlled trials. Eur J Clin Invest 2018;**48**:e12982.

18. BELLOU V, BELBASIS L, TZOULAKI I, EVANGELOU E, IOANNIDIS JP. Environmental risk factors and Parkinson's disease: An umbrella review of meta-analyses. Parkinsonism Relat Disord 2016;**23**:1-9.

19. WHO Collaborating Centre for Drug Statistics Methodology, ATC classification index with DDDs, 2019. Oslo, Norway. https://www.whocc.no/atc_ddd_index/. Accessed January 17, 2019.

20. SHEA BJ, REEVES BC, WELLS G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017;**358**:j4008.

21. DERSIMONIAN R, LAIRD N. Meta-analysis in clinical trials. Controlled clinical trials. 1986;**7**:177-88.

22. HIGGINS JPT, THOMPSON SG, SPIEGELHALTER DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc. 2009;**172**:137-59.

23. COCHRAN WG. The combination of estimates from different experiments. Biometrics. 1954;**10**:101-29.

24. STERNE JA, SUTTON AJ, IOANNIDIS JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011;**343**:d4002.

25. EGGER M, DAVEY SMITH G, SCHNEIDER M, MINDER C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;**315**:629-34.

26. IOANNIDIS JP, TRIKALINOS TA. An exploratory test for an excess of significant findings. Clin Trials 2007;**4**:245-53.

27. LUBIN JH, GAIL MH. On power and sample size for studying features of the relative odds of disease. Am J Epidemiol 1990;**131**:552-66..

28. TSILIDIS KK, KASIMIS JC, LOPEZ DS, NTZANI EE, IOANNIDIS JPA. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. BMJ 2015;**350**:g7607.

29. DRAGIOTI E, KARATHANOS V, GERDLE B, EVANGELOU E. Does psychotherapy work? An umbrella review of meta-analyses of randomized controlled trials. Acta psychiatr Scand 2017;**136**:236-46

30. MACHADO MO, VERONESE N, SANCHES M, et al. The association of depression and all-cause and cause-specific mortality: an umbrella review of systematic reviews and meta-analyses. BMC Med 2018;**16**:112.

31. BELBASIS L, BELLOU V, EVANGELOU E, IOANNIDIS JP, TZOULAKI I. Environmental risk factors and multiple sclerosis: an umbrella review of systematic reviews and meta-analyses. Lancet Neurol 2015;**14**:263-73.

32. GUYATT GH, OXMAN AD, VIST GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;**336**:924-6.

33. GUYATT GH, THORLUND K, OXMAN AD, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. J Clin Epidemiol 2013;**66**:173-83.

34. IORIO A, SPENCER FA, FALAVIGNA M, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. BMJ 2015;350.

35. BALSHEM H, HELFAND M, SCHUNEMANN HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;**64**:401-6.

36. PAPOLA D, OSTUZZI G, THABANE L, GUYATT G, BARBUI C. Antipsychotic drug exposure and risk of fracture: A systematic review and meta-analysis of observational studies. Int Clin Psychopharmacol 2018;**33**:181-96.

37. BARBUI C, CONTI V, CIPRIANI A. Antipsychotic drug exposure and risk of venous thromboembolism: A systematic review and meta-analysis of observational studies. Drug Saf 2014;**37**:79-90.

38. HSU WT, ESMAILY-FARD A, LAI CC, et al. Antipsychotics and the Risk of Cerebrovascular Accident: A Systematic Review and Meta-Analysis of Observational Studies. J Am Med Dir Assoc 2017;**18**:692-9.

39. HUANG KL, FANG CJ, HSU CC, WU SI, JUANG JJM, STEWART R. Myocardial infarction risk and antipsychotics use revisited: a meta-analysis of 10 observational studies. J Psychopharmacol 2017;**31**:1544-55.

40. NOSE M, RECLA E, TRIFIRO G, BARBUI C. Antipsychotic drug exposure and risk of pneumonia: A systematic review and meta-analysis of observational studies. Pharmacoepidemiol and Drug Saf 2015;**24**:812-20.

41. SALVO F, PARIENTE A, SHAKIR S, et al. Sudden cardiac and sudden unexpected death related to antipsychotics: A meta-analysis of observational studies. Clin Pharmacol Ther. 2016;**99**:306-14.

42. ZHANG Y, CHAPMAN A-M, PLESTED M, JACKSON D, PURROY F. The Incidence, Prevalence, and Mortality of Stroke in France, Germany, Italy, Spain, the UK, and the US: A Literature Review. Stroke Res Treat 2012;**2012**:436125-.

43. HEIT JA. Epidemiology of venous thromboembolism. Nat Rev Cardiol 2015;**12**:464-74.

44. TORRES A, PEETERMANS WE, VIEGI G, BLASI F. Risk factors for community-acquired pneumonia in adults in Europe: a literature review. Thorax 2013;**68**:1057-65.

45. ROGER VL. Epidemiology of myocardial infarction. Med Clin North Am 2007;**91**:537-52.

46. CORBETT A, BURNS A, BALLARD C. Don't use antipsychotics routinely to treat agitation and aggression in people with dementia. BMJ 2014;**349**:g6420.

47. HERMES EDA, SERNYAK M, ROSENHECK R. Use of second-generation antipsychotic agents for sleep and sedation: a provider survey. Sleep 2013;**36**:597-600.

48. KAMBLE P, CHEN H, SHERER JT, APARASU RR. Use of antipsychotics among elderly nursing home residents with dementia in the US: an analysis of National Survey Data. Drugs Aging. 2009;**26**:483-92.

49. ANDERSON FA, JR., SPENCER FA. Risk factors for venous thromboembolism. Circulation 2003;**107**:I9-16.

50. LEUCHT S, CIPRIANI A, SPINELI L, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. Lancet 2013;**382**:951-62.

51. BARBUI C. Sharing all types of clinical data and harmonizing journal standards. BMC Med 2016;**14**:63.

52. ROCKHOLD F, NISEN P, FREEMAN A. Data Sharing at a Crossroads. N Engl J Med 2016;**375**:1115-7.

53. MCDONALD L, SCHULTZE A, SIMPSON A, GRAHAM S, WASIAK R, RAMAGOPALAN SV. Lack of data sharing in observational studies. BMJ 2017;**359**:j4866.

Table 1. Main characteristics of studies included in the six meta-analyses

	First author	Year	Outcome	Outcome definition	N. of patients with the event	N. of patients without the event	Total N. of participants	Mean Age (SD)	Female sex (%)	Population†‡	Study design	Type of metric	ES	Lower 95% CI	Upper 95% CI
Individual studies included in the hip fracture review	Koponen	2017	HF	ICD-10	2971	54617	57588	64	-	Alzheimer disease	retrospective	HR	1.54	1.39	1.7
	Dennis	2017	HF	ICD-10	399	9035	9434	82	-	Dementia	retrospective	PERR	1.62	1.59	1.65
	Bakken	2016	HF	unclear	39938	866484	906422	73	56	Miscellaneous conditions	retrospective	SIR	2.1	1.9	2.2
	Leach	2015	HF	ICD-10	4418	4418	8828	88	63	Miscellaneous conditions	case-crossover	OR	1.47	1.21	1.8
	Bohlken	2015	HF	ICD-10	Unknown	Unknown	106312	81	61	Dementia	retrospective	HR	1.2	1.01	1.41
	Fraser	2015	HF	ICD-10	12315	183239	195554	81	65	Dementia	retrospective	OR	1.67	1.53	1.81
	Wu	2015	HF	ICD-9	605	2828	3433	58	45	Schizophrenia and related conditions	nested case-control	OR	1.34	1.07	1.69
	Rigler	2013	HF	unclear	79	8183	8262	83	72	Miscellaneous conditions (61% with dementia)	retrospective	HR	1.76	1.08	2.87
	Soresen	2013	HF	ICD-10	15431	3807597	3823028	48	-	Schizophrenia	case-control	RR	1.42	1.21	1.65
	Pouwels	2013	HF	unclear	275	9099	9374	74	42	Parkinson disease	retrospective	HR	1.24	0.79	1.96
	Baker	2010	HF	unclear	617	19487	20104	79	65	Dementia	retrospective	HR	3.6	1.3	10
	Howard	2007	HF	ICD-9	16341	29889	46230	79	79	Schizophrenia	case-control	OR	2.6	2.43	2.78
	Kaye	2004	HF	unclear	32827	65295	98122	48	55	Miscellaneous conditions	nested case-control	RR	1.4	1.2	1.6
	Jacquim-Gadda	1998	HF	self reported	285	2931	3216	75	-	Miscellaneous conditions	prospective	OR	1.41	0.46	4.3
	Jensen	1991	HF	ICD-9	200	200	400	81	-	Miscellaneous conditions	case-control	OR	1.47	0.73	2.97
	Jalbert	2010	HF	ICD-9	764	3582	4346	83	75	Miscellaneous conditions	nested case-control	OR	1.26	1.05	1.52
	Pouwels	2009	HF	unclear	6763	26341	33104	79	73	Miscellaneous conditions	case-control	OR	1.68	1.43	1.99
	Kolanowski	2006	HF	ICD-9	90	959	1049	75	-	Dementia	retrospective	OR	264	1.04	6.72
	Hugenholtz	2005	HF	ICD-9	22250	22250	44500	77	76	Miscellaneous conditions	case-control	OR	1.3	1.1	1.5

	First author	Year	Outcome	Outcome definition	N. of patients with the event	N. of patients without the event	Total N. of participants	Mean Age (SD)	Female sex (%)	Population†‡	Study design	Type of metric	ES	Lower 95% CI	Upper 95% CI
	Wang	2001	HF	unclear	1222	4888	6110	82	84	Miscellaneous conditions	case-control	OR	1.61	1.29	2.01
	Guo	1998	HF	ICD-9	134	1474	1608	82	-	Miscellaneous conditions	prospective	RR	0.86	0.4	1.84
	Lichtenstein	1994	HF	ICD-9	129	234	363	82	67	Miscellaneous conditions	case-control	OR	0.79	0.23	2.7
	Cumming	1993	HF	unclear	209	207	416	80	75	Miscellaneous conditions	case-control	OR	1.28	0.64	2.5
	Ray	1987	HF	ICD-9	1021	5606	6627	81	77	Miscellaneous conditions	case-control	OR	2	1.6	2.6
Hip fracture systematic review	**Papola[36]**	**2018**	**HF**		**159.283**	**5.128.843**	**5.394.430**	**55,4 (12.5)**	**52%**	**Dementia (6/24 studies) Schizophrenia (3/24) Parkinson/Parkinsonism (1/24) Miscellaneous conditions (14/24)**	**12 case-control, 2 prospective, 9 retrospective, 1 case cross-over**	**OR**	**1.57**	**1.42**	**1.74**
Individual studies included in the venous thromboembolism review	Allenet	2012	PE	ICD-9	76814	28646957	28723771	56	59	Miscellaneous conditions	retrospective	OR	1.17	1.13	1.21
	Hamanaka	2004	PE	Autopsy criteria	28	1097	1125	55	100	Miscellaneous conditions	case-control	OR	10.49	3.95	27.85
	Parkin	2003	PE	ICD-9	62	243	305	43	32	Miscellaneous conditions	case-control	OR	13.3	2.3	76.3
	Kleijer	2010	VTE	ICD-9	1032	4125	5157	76	67	Miscellaneous conditions	case-control	OR	0.9	0.73	1.1
	Ray	2002	VTE	ICD-9	22514	33033	55547	75	72	Miscellaneous conditions	retrospective	OR	1.1	0.95	1.27
	Liperoti	2005	VTE	ICD-9	539	131479	132018	80	72	Miscellaneous conditions	retrospective	HR	1.25	0.99	1.56
	Parker	2010	VTE	unclear	25532	89491	115023	67	56	Miscellaneous conditions	nested case-control	OR	1.32	1.19	1.47
	Hipplsley-Cox	2001	VTE	ICD-9	14756	2299945	2314701	47	49	Miscellaneous conditions	prospective	HR	1.67	1.41	1.98
	Jonsson	2009	VTE	ICD-8 and ICD-10	5999	59990	65989	65	55	Miscellaneous conditions	nested case-control	RR	1.87	1.53	2.28
	Ishiguro	2011	VTE	Autopsy criteria	112	375	487	40	-	Miscellaneous conditions	case-control	OR	2.11	0.68	6.75
	Masopust	2007	VTE	unclear	266	274	540	43	51	Miscellaneous conditions	case-control	OR	2.76	1.01	7.55
	Lacut	2007	VTE	unclear	677	677	1354	68	57	Miscellaneous conditions	case-control	OR	3.7	1.9	7.1

	First author	Year	Outcome	Outcome definition	N. of patients with the event	N. of patients without the event	Total N. of participants	Mean Age (SD)	Female sex (%)	Population†‡	Study design	Type of metric	ES	Lower 95% CI	Upper 95% CI
	Zornberg	2000	VTE	ICD-8	42	168	210	44	76	Miscellaneous conditions	case-control	OR	7.1	2.3	21.9
	Thomassen	2001	VTE	objective tests	474	474	948	43	-	Miscellaneous conditions	case-control	OR	9.07	0.48	169.05
Venous thromboembolism systematic review	**Barbui[37]**	**2014**	**VTE**		**148.847**	**31.268.328**	**31.417.175**	**55.5 (3.2)**	**58%**	**Miscellaneous conditions**	**10 case-control, 1 prospective, 3 retrospective**	**OR**	**1.55**	**1.31**	**1.83**
Individual studies included in the stroke review	Liperoti	2005	Stroke / TIA	ICD-9	1130	3658	4788	82	71	Dementia	case-control	OR	1.2	0.95	1.52
	Douglas	2008	Stroke	ICD-10	3395	3395	6790	80	-	Miscellaneous conditions	self-controlled	RR	1.73	1.6	1.87
	Kleijer	2009	Stroke	ICD-9	518	2030	2548	76	56	Miscellaneous conditions	nested case-control	OR	1.6	1.29	1.98
	Chan	2010	CVA	unclear	107	982	1089	80	66	Dementia	retrospective	HR	0.98	0.62	1.54
	Laredo	2011	CVA	ICD-9	3149	15613	18762	81	69	Dementia	nested case-control	OR	1.17	1.06	1.29
	Wang	2012	Stroke	ICD-9	255	256	511	69	2	Miscellaneous conditions	case-case time control	OR	1.8	1.7	1.9
	Hsieh	2013	Stroke / TIA	ICD-9	386	772	1158	57	50	Schizophrenia	nested case-control	OR	1.94	1.11	3.39
	Liu	2013	Stroke	ICD-9	811	1432	2243	78	53	Dementia	retrospective	HR	1.17	0.99	1.38
	Mundet-Tuduri	2013	Stroke	ICD-10	1084	26727	27811	54	53	Miscellaneous conditions	cross-sectional	OR	1.89	1.35	2.65
Stroke systematic review	**Hsu[38]**	**2017**	**Stroke**		**10.835**	**54.865**	**65.700**	**68,7 (13.8)**	**58%**	**Dementia (4/9 studies studies) Schizophrenia (1/9) Miscellaneous conditions (4/9)**	**4 case-control, 2 retrospective, 1 cross-sectional, 1 self-controlled case series, 1 case-case time control**	**OR**	**1.45**	**1.24**	**1.70**

	First author	Year	Outcome	Outcome definition	N. of patients with the event	N. of patients without the event	Total N. of participants	Mean Age (SD)	Female sex (%)	Population†‡	Study design	Type of metric	ES	Lower 95% CI	Upper 95% CI
Individual studies included in the myocardial infarction review	Thorogood	1992	Death by MI	ICD-9	161	309	470	27	100	Miscellaneous conditions (women)	case-control	RR	6.2	2	19.1
	Penttinen	1996	MI	ICD-9	83	249	332	-	0	Miscellaneous conditions (men)	case-control	OR	1.5	0.4	6
	Pratt	1996	MI	self-reported	63	1559	1622	40	62	Miscellaneous conditions	prospective	OR	2.92	1.23	6.98
	Enger	2004	MI	ICD-9	40	11480	11520	39	58	Schizophrenia	retrospective	RR	4.81	2.44	9.46
	Nakagawa	2006	MI	ICD-10	21377	106885	128262	69	39	Miscellaneous conditions	case-control	RR	0.99	0.96	1.02
	Lin	2014	MI	ICD-9	29903	29903	59806	71	48	Schizophrenia (10%); Mood disorders (36%); Dementia (54%).	case-crossover	OR	2.52	2.37	2.68
	Brauer	2015	MI	unclear	734	734	1468	70	44	Miscellaneous conditions	self-controlled case series	RR	2.76	2.02	3.77
	Wu	2015	MI	ICD-9	295	296	834	57	40	Schizophrenia	case-crossover	OR	2.13	1.49	3.06
	Hwang	2014	MI	ICD-9	97777	97777	195554	81	-	Miscellaneous conditions	retrospective	RR	1.13	0.97	1.32
Myocardial infarction systematic review	**Huang[39]**	**2017**	**MI**		**150.433**	**249.192**	**399.625**	**74,1 (9.2)**	**46%**	**Schizophrenia (2/9 studies) Miscellaneous population (7/9)**	**3 case-control, 1 prospective, 2 retrospective, 2 case cross-over, 1 self-controlled case series**	**OR**	**2.21**	**1.41**	**3.46**
Individual studies included in the pneumonia review	Gau	2010	PN	Medical records and radiographic findings	194	952	1146	80	64	Miscellaneous conditions	case-control	OR	2.26	1.23	4.15
	Knol	2008	PN	Medical records	543	2163	2706	80	64	Miscellaneous conditions	nested case-control	OR	1.9	0.5	7.4
	Kuo	2013	PN	ICD	1739	6949	8688	63	37	Schizophrenia	case-control	RR	1.64	1.4	1.92

	First author	Year	Outcome	Outcome definition	N. of patients with the event	N. of patients without the event	Total N. of participants	Mean Age (SD)	Female sex (%)	Population†‡	Study design	Type of metric	ES	Lower 95% CI	Upper 95% CI
	Pratt	2011	PN	ICD	1914	11410	13324	85	-	Miscellaneous conditions	self-controlled case series	IRR	1.64	1.32	2.04
	Trifirò	2010	PN	Medical records and radiographic findings	258	1689	1947	83	71	Miscellaneous conditions	nested case-control	OR	1.97	1.45	2.69
	Yang	2013	PN	ICD	258	657	915	44	39	Bipolar Disorder	case-control	RR	2.19	1.8	2.65
Pneumonia systematic review	**Nosé[40]**	**2015**	**PN**		**4.906**	**23.820**	**28.726**	**76,2 (12.3)**	**49%**	**Schizophrenia (1/6 studies) Bipolar disorder (1/6 studies) Miscellaneous conditions (4/6)**	**5 case-control 1 self-controlled case series**	**OR**	**1.84**	**1.62**	**2.09**
Individual studies included in the sudden cardiac death review	Jones	2013	life-threatening VA	ICD-10	1146	376166	377312	56	58	Psychiatric disorders	retrospective	RR	1.16	1.02	1.31
	Ray	2009	SCD by VA	Deaths compatible with SCD	1684	278216	279900	46	67	Miscellaneous conditions	retrospective	IRR	1.91	1.65	2.21
	Jolly	2009	SCD by VA	post-mortem examination	1010	3030	4040	67	33	Miscellaneous conditions	case-control	OR	4.29	2.68	6.88
	Van Noord	2011	SCD by VA	Deaths compatible with SCD	1424	14443	15867	70	40	Miscellaneous conditions	case-control	OR	3.9	2.06	7.37
	Reilly	2002	SCD by VA	Deaths compatible with SCD	69	138	207	66	67	Psychiatric disorders	case-control	OR	7.6	0.54	108.1
	Kenbubpha	2002	SCD by VA	Deaths compatibl	54	108	162	40	37	Miscellaneous conditions	case-control	OR	1.58	0.68	3.65

	First author	Year	Outcome	Outcome definition	N. of patients with the event	N. of patients without the event	Total N. of participants	Mean Age (SD)	Female sex (%)	Population†‡	Study design	Type of metric	ES	Lower 95% CI	Upper 95% CI
				e with SCD											
Sudden cardiac death systematic review	**Salvo[41]**	**2016**	**SCD by VA**		**5·387**	**672.101**	**677.488**	**52.2 (6.2)**	**61%**	**Psychiatric disorders (2/6 studies) Miscellaneous conditions (4/6)**	**4 case-control, 2 retrospective**	**OR**	**2.24**	**1.45**	**3.46**

Explanations

ICD = International Classification of Diseases; OR = Odds Ratio; RR = Relative Risk; IRR = International Rate Ratio; PERR = Prior Event Risk Ratio; SIR = Standardized Incidence Ratio; ES = Effect Size; HF= Hip Fracture; PE = Pulmonary Embolism; VTE = Venous Thromboembolism; CVA = cerebrovascular accident; MI = Myocardial Infarction; TIA= Transient Ischemic Attack; PN = Pneumonia; SCD = Sudden Cardiac Death.

† "Miscellaneous population": it is likely that patients considered for these studies suffered from medical/psychiatric condition that justified an antipsychotic prescription, but the original investigation didn't report it or did it in in a way that cannot be used to identify a well-defined population.

‡ mean age of participants ordered by diagnosis (SD): HIP FRACTURE: dementia: 78,4 (11.5) years; schizophrenia: 48,38 (5.7) years; Parkinson: 74 years; miscellaneous population: 71,5 (9.8) years. VENOUS THROMBOEMBOLISM: miscellaneous population: 55,5 (7.2) years. STROKE: dementia: 80,9 (10.3) years; miscellaneous population: 60,4 (10.6) years. MYOCARDIAL INFARCTION: miscellaneous population: 75,2 (13.1) years; psychiatric population: 40,2 (5.5) years. PNEUMONIA: miscellaneous population: 83,8 (11.7) years; psychiatric population: 61,2 (8.2) years. SUDDEN CARDIAC DEATH: miscellaneous population: 47,5 (5.9) years; psychiatric population: 56 (6.2) years.

Table 2. AMSTAR-2 appraisal

		Papola et al., 2018[36] - Hip fracture	**Barbui et al., 2014[37] - VTE**	**Hsu et al., 2017[38] - stroke**	**Haung et al., 2017[39] - MI**	**Nosé et al., 2015[40] - Pneumonia**	**Salvo et al., 2016[41] - SCD**
1	Did the research questions and inclusion criteria for the review include the components of PICO?	√	√	√	√	√	√
2	**Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?**	√	√	×	×	√	partially
3	Did the review authors explain their selection of the study designs for inclusion in the review?	√	√	×	√	√	√
4	**Did the review authors use a comprehensive literature search strategy?**	√	√	√	√	√	√
5	Did the review authors perform study selection in duplicate?	√	√	√	√	√	×
6	Did the review authors perform data extraction in duplicate?	√	√	×	√	√	×
7	**Did the review authors provide a list of excluded studies and justify the exclusions?**	√	partially	partially	×	×	√
8	Did the review authors describe the included studies in adequate detail?	√	√	partially	√	√	√
9	**Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?**	√	√	√	√	√	√
10	Did the review authors report on the sources of funding for the studies included in the review?	×	×	×	×	×	×
11	**If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?**	√	√	√	√	√	√

12	If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	√	√	√	×	√	√
13	**Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?**	√	√	√	×	√	√
14	Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	√	√	√	√	√	√
15	**If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?**	√	×	√	√	√	√
16	Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	√	√	×	√	√	√
Reporting quality of each meta-analysis		High	low	low	Critically low	low	moderate

Explanation
"Critical" domains are in bold.

Table 3. Strength of association according to umbrella review criteria

	Features used for classification of associations according to Umbrella Review Criteria									
Outcome	Number of cases	Random-effects P-value	I^2	MA Predictive Intervals	Random-effects ES (95% CI) of the largest study	Egger's test P-value	Significant studies			**Strength of association**
							Observed	Expected§	P-value†	
Pneumonia	4.906	$7{,}1 \times 10^{-21}$	27	1.38 – 2.45	1.64 (1.40 – 1.92)	0.35	4	6	NR	**I**
Hip Fracture	159.283	1.2×10^{-18}	92	1.03 – 2.41	1.62 (1.59 – 1.65)	0.76	10	17.2	NR	**II**
Venous Thromboembolism	148.847	3.4×10^{-7}	87	0.91 – 2.65	1.17 (1.13 – 1.21)	0.05	3	12.2	NR	**II**
Stroke	10.835	$2{,}9 \times 10^{-6}$	91	0.86 – 2.45	1.80 (1.70 – 1.90)	0.65	4	7.34	NR	**III**
Myocardial Infarction	150.433	0.00053	99	0.47 – 10.42	0.99 (0.96 – 1.02)	0.16	1	6.53	NR	**III**
Sudden Cardiac Death	5.387	0.00029	91	0.56 – 8.95	1.16 (1.02 – 1.31)	0.19	2	5.11	NR	**III**

Explanations

ES = effect-size; MA = meta-analysis; NR = not relevant.

§ Observed and expected number of significant studies using effect of largest study (smallest SE) of each meta-analysis as plausible effect size.

†P value of excess significance test. All statistical tests two sided.

Table 4. GRADE appraisal

Certainty assessment								Relative (95% CI)	Certainty
Outcome	**№ of studies**	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**		
Hip fracture	24	observational studies	not serious	not serious	not serious	not serious	none	**OR 1.57** (1.42 to 1.74)	⊕⊕◯◯ LOW
Thromboembolism	14	observational studies	not serious	serious [d]	not serious	not serious	publication bias strongly suspected	**OR 1.55** (1.31 to 1.83)	⊕◯◯◯ VERY LOW
Stroke	9	observational studies	not serious	serious [c]	not serious	not serious	none	**OR 1.45** (1.24 to 1.70)	⊕◯◯◯ VERY LOW
Myocardial Infarction	9	observational studies	not serious	serious [a]	not serious	serious [b]	publication bias strongly suspected;	**OR 2.21** (1.41 to 3.46)	⊕◯◯◯ VERY LOW
Pneumonia	6	observational studies	not serious	not serious	not serious	not serious	none	**OR 1.84** (1.62 to 2.09)	⊕⊕◯◯ LOW

Certainty assessment								Relative (95% CI)	Certainty
Outcome	**№ of studies**	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**		
Sudden cardiac death	6	observational studies	not serious	not serious	not serious	serious [b]	none	**OR 2.24** (1.45 to 3.47)	⨁◯◯◯ VERY LOW

CI: Confidence interval; OR: Odds ratio

Explanations

a. Visual inspection of forest plot suggested heterogeneity; I-squared = 99%;
b. Wide confidence interval;
c. Visual inspection of forest plot suggested heterogeneity; I-squared = 91%;
d. Visual inspection of forest plot suggested heterogeneity; I-squared = 87%.

FIGURE LEGENDS

Figure 1. PRISMA flow-diagram.

Figure 2. Ranking of associations on the risk of life-threatening medical events associated with exposure to antipsychotic drugs

Explanations:

[1] High certainty: we are very confident that the true effect lies close to that of the estimate of the effect. Moderate certainty: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different. Low certainty: The true effect may be substantially different from the estimate of the effect. Very low certainty: the true effect is likely to be substantially different from the estimate of effect.

[2] Convincing evidence (Class I): more than 1000 cases + significant summary associations (p<10-6) per random-effects calculations + no evidence of small-study effects + no evidence of excess of significance bias + prediction intervals not including the null value + largest study nominally significant (p<0.05) + not large heterogeneity (i.e., I2< 50%). Highly Suggestive evidence (Class II): more than 1000 cases + significant summary associations (p<10-6) per random-effects calculation + largest study nominally significant (p<0.05). Suggestive Evidence (Class III): more than 1000 cases + significant summary associations (p<10-3) per random-effects calculations. Weak evidence (Class IV): All other associations with p < 0.05. Non-significant associations: All associations with p >0.05.

[3] High: No or one non-critical weakness: the systematic review provides an accurate and comprehensive summary of the results of the available studies that address the question of interest. Moderate: more than one non-critical weakness: the systematic review has more than one weakness but no critical flaws. It may provide an accurate summary of the results of the available studies that were included in the review. Low: one critical flaw with or without non-critical weaknesses: the review has a critical flaw and may not provide an accurate and comprehensive summary of the available studies that address the question of interest. Critically low: more than one critical flaw with or without non-critical weaknesses: the review has more than one critical flaw and should not be relied on to provide an accurate and comprehensive summary of the available studies.